

Fig. 1

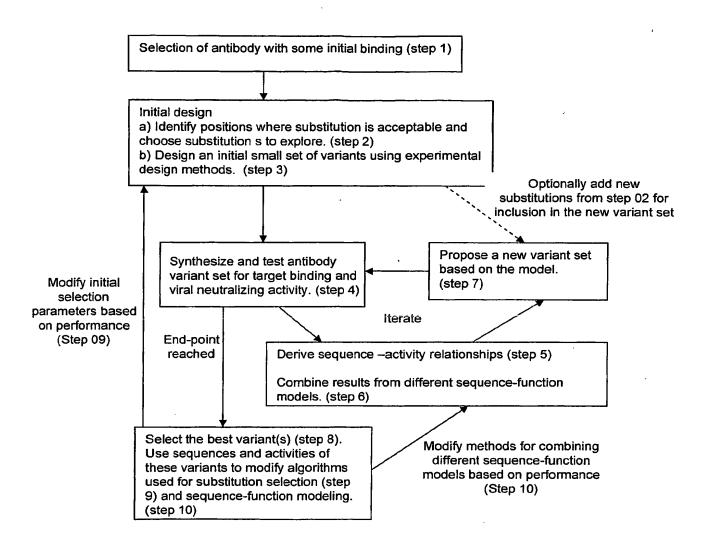


Figure 2

Select initial candidate antibody sequence(s)

A: Substitutions from homologous sequences in framework & CDR

- -Identify framework sequences within a defined evolutionary distance (PAM units)
- Reconstruct phylogenetic tree for framework region only

RULE 1a:

- Select a defined number of framework residues that have undergone advantageous change RULE 2a:
- Select defined number of framework and defined number of CDR positions with highest mutability index

RULE 3a:

 Select <u>all</u> amino acid substitutions from sequences in the same Chothia class SCORE:: Weighted value for each rule satisfied

B: Substitutions from homologous structures

Superpose homologous structures from PDB

RULE 1b:

- Calculate ∆G change for all single substitutions
- Select changes with <a defined value Kcal/mol change in free energy

RULE 2b:

- Estimate mean RMSD for every window of a defined number of residues
- Select framework sites with an RMSD greater than a defined value
- · Identify changes found in homologous sequences
- Select framework varying sites within a defined distance from CDR

SCORE:: Weighted value for each rule satisfied

C: Substitutions from substitution matrices

 Calculate substitution matrix for all framework regions and canonical classes, rank all possible single substitutions for favorability

RULE 1c:

· Select highest scoring substitutions for each matrix

RULE 2c:

- Rank all possible single substitutions for favorability using a universal substitution matrix
- Select highest scoring positions

SCORE: Weighted value for each rule satisfied

D: Substitutions from PCA analysis

- Determine principal components of sequence variation in alignment of homologs
 RULE 1d:
- Group CDRs based on PCA of amino acid sequences in the CDR.
- Select highest scoring CDR positions that differentiate antibody sequences by function RULE 2d:
- · Group CDR positions based on observed amino acid frequencies at each site
- Rank the groups based on contributions to one or more principal components
- · Select the top groups of sites to vary.

SCORE: Weighted value for each rule satisfied

E: Substitutions from Binding pocket Analysis

RULE 1e:

 Select sites based where physico-chemical properties of residues are conserved in the pocket

RULE 2e:

- Select CDR changes derived from evolutionary models to correlate properties of amino acids
- SCORE: Weighted value for each rule satisfied

Figure 3

- A: Calculate average weight vectors:
- a) Build sequence-function model from data in which rows (sequence + function) and/or columns (substitutions) are randomly omitted. Calculate weight vectors.
- b) Repeat 1,000 times.
- c) Calculate average values and standard deviations for weight vectors and rank in order of importance.

Figure 4

- A: Calculate average weight vectors:
- a) Build sequence-function model from data in which rows (sequence + function) and/or columns (substitutions) are randomly omitted. Calculate weight vectors.
- b) Repeat 1,000 times.
- c) Calculate average values and standard deviations for weight vectors and rank in order of importance.

- B: Calculate weight vectors from randomized data;
- a) Randomly associate sequence data with function data
- b)Build sequence-function model and calculate weight vectors.
- c) Repeat 1,000 times.
- d) Calculate average value and standard deviations for weight vectors from randomized data.

C: Calculate number of standard deviations weight vector is above value from randomized data.

Figure 5

```
E coli leader peptide
          -10
MKKLLFAIPL VVPFYSHSTM (SEQ ID NO.: 1)
Proteinase K
                                 31
                      21
           11
APAVEORSEA APLIEARGEM VANKYIVKFK EGSALSALDA AMEKISGKPD
                      71
                                 81
HVYKNVFSGF AATLDENMVR VLRAHPDVEY IEQDAVVTIN AAQTNAPWGL
                                 131
                      121
           111
ARISSTSPGT STYYYDESAG QGSCVYVIDT GIEASHPEFE GRAQMVKTYY
                                 181
YSSRDGNGHG THCAGTVGSR TYGVAKKTQL FGVKVLDDNG SGQYSTIIAG
         . 211
                                 231
                      221
MDFVASDKNN RNCPKGVVAS LSLGGGYSSS VNSAAARLQS SGVMVAVAAG
                      271
                                 281
NNNADARNYS PASEPSVCTV GASDRYDRRS SFSNYGSVLD IFGPGTSILS
                                 331
           311
                      321
TWIGGSTRSI SGTSMATPHV AGLAAYLMTL GKTTAASACR YIADTANKGD
           361
                      371
LSNIPFGTVN LLAYNNYQAV DHHHHHH (SEQ ID NO.: 2)
```

Figure 6

-60	-50	-40	-30	-20	-10 -1
atgaaaaaac	tgctgttcgc	gattccgctg	gtggtgccgt	tctatagcca	tagcaccatg
1	11	21	31	41	51
GCACCGGCCG	TTGAACAGCG	TTCTGAAGCA	GCTCCTCTGA	TTGAGGCACG	TGGTGAAATG
61	71	81	91	101	111
GTAGCAAACA	AGTACATCGT	GAAGTTCAAG	GAGGGTTCTG	CTCTGTCTGC	TCTGGATGCT
121	131	141	151	161	171
GCTATGGAAA	AGATCTCTGG	CAAGCCTGAT	CACGTCTATA	AGAACGTGTT	CAGCGGTTTC
181	191	201	211	221	231
GCAGCAACTC	TGGACGAGAA	CATGGTCCGT	GTACTGCGTG	CTCATCCAGA	CGTTGAATAC
241	251	261	271	281	291
ATCGAACAGG	ACGCTGTGGT	TACTATCAAC	GCGGCACAGA	CTAACGCACC	TTGGGGTCTG
301	311	321	331	341	351
GCACGTATTT	CTTCTACTTC	CCCGGGTACG	TCTACTTACT	ACTACGACGA	GTCTGCCGGT
361	371	381	391	401	411
CAAGGTTCTT	GCGTTTACGT	GATCGATACG	GGCATCGAGG	CTTCTCATCC	TGAGTTTGAA
421	431	441	451	461	471
GGCCGTGCAC	AAATGGTGAA	GACCTACTAC	TACTCTTCCC	GTGACGGTAA	TGGTCACGGT
481	491	501	511	521	531
ACTCATTGCG	CAGGTACTGT	TGGTAGCCGT	ACCTACGGTG	TTGCTAAGAA	AACGCAACTG
541	551	561	571	581	591
TTCGGCGTTA	AAGTGCTGGA	CGACAACGGT	TCTGGTCAGT	ACTCCACCAT	TATCGCGGGT
601	611	621	631	641	651
ATGGATTTCG	TAGCGAGCGA	TAAAAACAAC	CGCAACTGCC	CGAAAGGTGT	TGTGGCTTCT
661	671	681	691	701	711
CTGTCTCTGG	GTGGTGGTTA	CTCCTCTTCT	GTTAACAGCG	CAGCTGCACG	TCTGCAATCT
721	731	741	751	761	771
TCCGGTGTCA	TGGTCGCAGT	AGCAGCTGGT	AACAATAACG	CTGATGCACG	CAACTACTCT
781	791	801	811	821	831
CCTGCTAGCG	AGCCTTCTGT	TTGCACCGTG	GGTGCATCTG	ATCGTTATGA	TCGTCGTAGC
841	851	861	871	881	891
TCCTTCAGCA	ACTATGGTTC	CGTCCTGGAT	ATCTTCGGCC	CTGGTACTTC	TATCCTGTCT

Figure 7A

901 ACCTGGATTG	911 GCGGTAGCAC	921 TCGTTCCATT	931 TCCGGTACGA	941 GCATGGCTAC	951 TCCACATGTT
	971 CAGCATACCT		991 GGTAAGACCA	1001 CTGCTGCATC	1011 CGCTTGTCGT
	1031 ATACTGCGAA	1041 CAAAGGCGAT		1061 TCCCGTTCGG	1071 CACCGTTAAT
	1091 ACAACAACTA			1121 atcatcatca	1131 tag
(SEQ ID NO	.: 3)				

Figure 7B

```
gi | 19171215 | emb | CAD20578.1 | /89
gi | 19171217 | emb | CAD20579.1 | /1-
gi | 19171219 | emb | CAD20580.1 | /1-
gi | 19171221 | emb | CAD20581.1 | /1-
gi|16215662|emb|CAC95042.1|/90
gi|16506136|dbj|BAB70705.1|/78
gi | 16506134 | dbj | BAB70704.1 | /78
gi|16506140|dbj|BAB70707.1|/78
gi|16215677|emb|CAC95049.1|/90
gi | 117631 | sp | P29138 | CUDP_METAN
gi|6624958|emb|CAB63911.1|/90-
gi|16215669|emb|CAC95045.1|/90
gi|460032|gb|AAA91584.1|/84-36
gi | 6634475 | emb | CAB64346.1 | /87-
gi|16215664|emb|CAC95043.1|/87
g1 | 2351388 | gb | AAC49831.1 | /86-3
gi | 8671180 | emb | CAB95012.1 | /85-
gi | 16215666 | emb | CAC95044.1 | /85
gi | 16215671 | emb | CAC95046.1 | /85
gi | 4092486 | gb | AAC99421.1 | /64-2
gi | 18542429 | gb | AAL75579.1 | AF46
SUTIKA/91-367
gi | 131077 | sp | P06873 | PRTK_TRIAL
gi 230675 pdb 2PRK /1-277
gi | 494434 | pdb | 1PEK | E/1-277
gi|224977|prf||1205229A/1-275
gi|14278658|pdb|11C6|A/1-277
gi | 131084 | sp | P23653 | PRTR_TRIAL
gi|4761119|gb|AAD29255.1|AF104
gi|14626933|gb|AAK70804.1|/81-
gi|639712|gb|AAC48979.1|/83-34
gi|742825|prf||2011184A/84-362
gi|628051|pir||JC2142/84-362
gi|15808791|gb|AAL08502.1|AF41
gi|15808805|gb|AAL08509.1|AF41
gi|28918475|gb|EAA28148.1|/90-
gi|10181226|gb|AAC27316.2|/92-
gi|131088|sp|P20015|PRTT_TRIAL
gi|9971109|emb|CAC07219.1|/86-
gi | 7543916 | emb | CAB87194.1 | /89-
gi|5813790|gb|AAD52013.1|AF082
gi|23894244|emb|CAD23614.1|/11
gi|22652141|gb|AAN03634.1|AF40
gi|24528136|emb|CAD24010.1|/10
gi|24528132|emb|CAD24008.1|/10
A35742./126-403
gi|114081|sp|P08594|AQL1_THEAQ
AAA82980./129-408
gi | 15640187 | ref | NP_229814.1 | /1
AAA22247./107-381
```

Figure 8

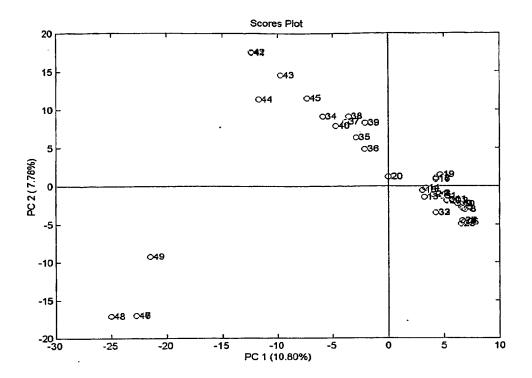


Figure 9

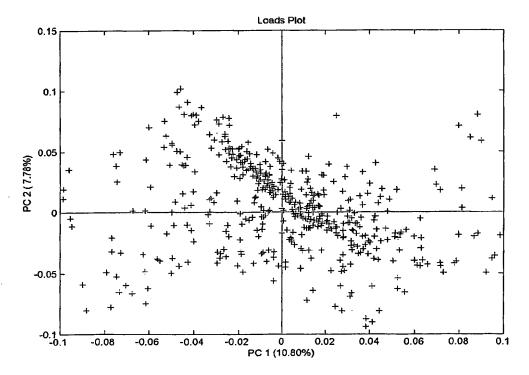


Figure 10

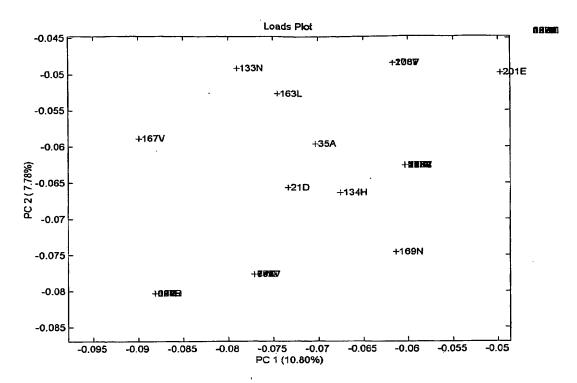


Figure 11

Residue	PC1 contrib.	PC2 contrib.	PC1+2 contrib.
15D	-0.0881	-0.0803	-0.1684
18D	-0.0881	-0.0803	-0.1684
19Q	-0.0881	-0.0803	-0.1684
22L	-0.0881	-0.0803	-0.1684
23P	-0.0881	-0.0803	-0.1684
65Y	-0.0881	-0.0803	-0.1684
66D	-0.0881	-0.0803	-0.1684
110R	-0.0881	-0.0803	-0.1684
137P	-0.0881	-0.0803	-0.1684
164D	-0.0881	-0.0803	-0.1684
189C	-0.0881	-0.0803	-0.1684
198R	-0.0881	-0.0803	-0.1684
8P	-0.0772	-0.0777	-0.1549
34T	-0.0772	-0.0777	-0.1549
67A	-0.0772	-0.0777	-0.1549
75Q	-0.0772	-0.0777	-0.1549
161T	-0.0772	-0.0777	-0.1549
199V	-0.0772	-0.0777	-0.1549
167V	-0.0899	-0.0589	-0.1488
21D	-0.0733	-0.0657	-0.1390
169N	-0.0613	-0.0746	-0.1359
134H	-0.0675	-0.0664	-0.1339

Figure 12

Variation	Score	Primary contribution to score
N95C	5	Structural stability at higher temperature: from published literature
P97S	3	P to S for flexibility and structural perturbabtion
S107D	5	from active homologs
S123A	7	Thermostable consensus
E138A	5	From experiments in literature
M145F	5	From experiments to improve thermostability
Y151A	8	From experiments to improve thermostability
V167 I	10	Allow user specified conservative changes (controlled perturbation)
L180I	10	Allow user specified conservative changes (controlled perturbation)
Y194S	10	Varaiation observed in highly active clone from our initial exp.
A199S	8	Allow user specified conservative changes (controlled perturbation)
K208H	7	PCA modelling of homologs collected from GenBank.
A236V	7	PCA modelling of homologs collected from GenBank.
R237N	5	From experiments to improve thermostability (in literature)
P265S	3	P to S for flexibility and structural perturbabtion
V267I	10	Allow user specified conservative changes (controlled perturbation)
S273T	15	Multiple sources identify this change. (thermostability and other)
G293A	8	For thermostability considerations (observed in thermitases)
L299C	5	For disulphide bridges with N95C (from literature)
I310K	5	from structural studies
K332R	8	for thermostability considerations (observed in thermitases)
S337N	8	for thermostability considerations (observed in thermitases)
P355S	3	P to S for flexibility and structural perturbabtion

Figure 13

variant-1: 123, 151, 293, 310, 332, 355 variant-2: 95, 145, 167, 199, 237, 273 variant-3: 97, 138, 180, 194, 236, 267 variant-4: 107, 132, 208, 265, 299, 337 variant-5: 123, 145, 151, 167, 273, 337 variant-6: 97, 107, 180, 236, 237, 310 variant-7: 123, 138, 199, 208, 265, 355 variant-8: 95, 194, 267, 293, 299, 332 variant-9: 95, 132, 138, 145, 167, 208 variant-10: 236, 237, 273, 293, 332, 355 variant-11: 97, 123, 265, 299, 310, 337 variant-12: 107, 151, 180, 194, 199, 267 variant-13: 95, 107, 123, 180, 194, 337 variant-14: 138, 151, 167, 199, 208, 299 variant-15: 97, 145, 237, 273, 293, 310 variant-16: 132, 236, 265, 267, 332, 355 variant-17: 97, 151, 199, 236, 299, 355 variant-18: 95, 107, 167, 180, 293, 310 variant-19: 145, 237, 265, 267, 332, 337 variant-20: 123, 132, 138, 194, 208, 273 variant-21: 123, 208, 236, 267, 293, 299 variant-22: 107, 132, 138, 145, 337, 355 variant-23: 97, 180, 194, 199, 265, 310 variant-24: 95, 151, 167, 237, 273, 332

Figure 14

Variant #	Changes	Reasons
variant-25	95	Confirm detrimental effect on enzyme
variant-26	97	Confirm detrimental effect on enzyme
variant-27	138	Confirm detrimental effect on enzyme
variant-28	208	Confirm detrimental effect on enzyme
variant-29	236	Confirm detrimental effect on enzyme
variant-30	237	Confirm detrimental effect on enzyme
variant-31	265	Confirm detrimental effect on enzyme
variant-32	299	Confirm detrimental effect on enzyme
variant-33	107, 123, 145	New combinations of positive changes
variant-34	151, 167, 180	New combinations of positive changes
variant-35	194, 199, 267	New combinations of positive changes
variant-36	273, 293, 310	New combinations of positive changes
variant-37	332, 337, 355	New combinations of positive changes
variant-38	107, 151, 194, 273, 332	New combinations of positive changes
variant-39	123, 167, 199, 293, 337	New combinations of positive changes
variant-40	145, 180, 267, 310, 355	New combinations of positive changes
variant-41	107, 167, 267, 273, 337	New combinations of positive changes
variant-42	123, 180, 194, 293, 355	New combinations of positive changes
variant-43	145, 151, 199, 310, 332	New combinations of positive changes
variant-44	145, 167, 194	New combinations of positive changes
variant-45	180, 199, 273	New combinations of positive changes
variant-46	267, 293, 332	New combinations of positive changes
variant-47	310, 337, 107	New combinations of positive changes
variant-48	355, 123, 151	New combinations of positive changes

Figure 15

Sequen	ce cha	nges	<u> </u>							v	ariar	ıts							
Position	WT	Mut	10	12	13	14	15	19	20	21	22	23	24	25	26	27	29	30	31
25	Y	H					<u> </u>	<u> </u>		 			Н		 _ `	 -		150	+
34	Α	S							1	\vdash			 -	 	S	 	\vdash	 	
48	K	E							f	 				 	-	\vdash	 		
50	D	N						-	†	N					<u> </u>		<u> </u>	 	 -
55	N	S							t —				†	s	 	t^-	 -	┼──	\vdash
63	Т	S										 		Ť	 	 	\vdash	 	
88	T	I						<u> </u>		l —	 				 	 	 	 	\vdash
95	N	C						_	1						 		 	├─	
97	P	S							-	 		_		<u> </u>		 	\vdash	 	
107	S	D				D			 			-					 	D	D
123	S	Α		A				\vdash	\vdash			_			<u> </u>	 		1	1-
132	I	V			V	-			<u> </u>						t	<u>-</u>	 	 	
138	E	Α							T									 	
145	M	F			F				 -									<u> </u>	
151	Y	Α		Α	Α	Α						A	A	Α				A	A
167	V	I			I							I	I	I			<u> </u>		┝∸
180	L	I				I						Ī	I	Ī		├ ─	<u> </u>		
194	Y	S	S			S									S		†	S	S
199	Α	S				S									S			-7-	
208	K	Н						Н							<u> </u>				
209	N	K									-								
233	S	N				•													
236	Α	V																	
237	R	N					N		N	N									<u> </u>
265	P	S									S								
267	V	I				I									I				
273	S	T			T		_T									T		Т	T
293	G	A		Α			Α									Α			
299	L	С																	
310	I	K		K												K			
332	K	R		R			R										R	R	R
337	S	N			N									i			N		
355	P	S		S			S										S		
362	L	М							М										
363	Α	V										V	-1						
369	A	V																	$\neg \neg$

Figure 16A

Sequen	ce cha		L							V	aria	ıts							
Position	WT	Mut	32	33	35	36	37	38	39	40	41	42	43	45	46	47	48	49	50
25	Y	H																	
34	Α	S		L											Γ		<u> </u>		
48	K	E		<u> </u>															
50	D	N																	
55	N	S	<u> </u>	L															
63	T	S																	
88	Т	I														I			
95	N	С																	
97	P	S																	
107	S	D			D							D							
123	S	Α	Α	Α		Α	A						Α						_
132	I	V																	
138	E	Α																	_
145	M	F						F											
151	Y	Α											A			Α	Α	Α	Α
167	V	Ĭ	I	I	I			I								I	I	I	Ī
180	L	Ī				I	I		Ī				·			I	Ī	Ī	I
194	Y	S				S	S	S											
199	Α	S	S	S		-			S										
208	K	Н																	
209	N	K										*							
233	S	N																	
236	A	V																	
237	R	N												N	N				
265	P	S																	
267	V	I			I					1	I					$\neg \neg$			
273	S	T			T			İ	T										
293	G	Α	Α	Α		Α	Α			A	Α								
299	L	С																	
310	I	K										K							
332	K	R								R	R								
337	S	N	N	N	N						Ī	N							
355	P	S				S	S						S						
362	L	M										\neg	$\neg \neg$		一十			_	
363	A	V						V											
369	Α	V										T		$\overline{\mathbf{v}}$					

Figure 16B

Sequen	ce chai	nges	T					V	arian	ts				
Position	WT	Mut	N2	N3	N4	N6	N7	N8	N9	N10	N11	NI3	N14	N15
25	Y	Н												
34	Α	S												
48	K	Е											<u> </u>	
50	D	N												
55	N	S												
63	T	S										T		
88	T	I												
95	N	С	С					С	С			С		<u> </u>
97	P	S		S		S					S	1	1	S
107	S	D			D	D						D		
123	S	Α					Α				Α	Α		
132	I	V			V				V		<u> </u>	l		
138	Е	Α		Α			Α		Α				Α	
145	M	F	F						F	[<u> </u>	<u> </u>	F
151	Y	Α								L	<u> </u>	<u> </u>	A	<u> </u>
167	V	Ī	I						I		<u> </u>	<u> </u>	I	
180	L	I		I		I					<u> </u>	1		<u> </u>
194	Y	S		S			<u> </u>	S	<u> </u>	<u> </u>		S	<u> </u>	
199	A	S	S				S			<u> </u>		<u> </u>	S	
208	K	Н			H		H		H			<u> </u>	Н	<u> </u>
209	N	K											1	<u> </u>
233	S	N			<u> </u>		<u>L.</u>		<u> </u>					<u> </u>
236	Α	V		V	<u> </u>	V			ļ	V			l	
237	R	N	N	<u> </u>	<u> </u>	N	<u> </u>			N				N
265	P	S	<u> </u>	<u> </u>	S	<u> </u>	S		L		S	ļ	ļ	<u> </u>
267	V	I	<u> </u>	I	<u> </u>	<u> </u>	<u> </u>	I	<u> </u>	L	<u> </u>	<u> </u>	-	<u> </u>
273	S	T	T	<u> </u>		<u> </u>	<u> </u>		<u> </u>	T		ļ	<u> </u>	T
293	G	A	<u> </u>	<u> </u>		<u> </u>	ļ	A	ļ	A	<u> </u>	↓		Α
299	L	С		<u> </u>	C	1	<u> </u>	C	 	ļ	С	ļ	С	<u> </u>
310	I	K	<u> </u>	<u> </u>	ļ	K	<u> </u>	ļ	↓	<u> </u>	K	L	 	K
332	K	R		<u> </u>	<u> </u>	<u> </u>	<u> </u>	R	<u> </u>	R	l		ļ	
337	S	N	<u> </u>	<u> </u>	N_	 	 _			<u> </u>	N	N	1	<u> </u>
355	P	S		<u> </u>		<u> </u>	S		<u> </u>	S	 	ļ	ļ	
362	L	M		ļ	ļ	 -			<u> </u>	ļ	ļ	_	ļ	
363	Α	V			<u> </u>	<u> </u>		<u> </u>	<u> </u>		<u> </u>	 		_
369	A	V	<u></u>	<u> </u>	<u> </u>	<u>L</u>	<u></u>	<u> </u>	<u> </u>	<u> </u>	L	<u> </u>	<u> </u>	<u> </u>

Figure 16C

Sequence	e char	nges								V	arian	ts							
Position	WT	Mut	N16	N17	N18	N19	N20	N21	N22	N23	N24	N25	N26	N27	N28	N30	N33	N40	N43
25	Y	H																	
34	Α	S										Ĺ							
48	K	E																	
50	D	N															L		L
55	N	S																L	
63	Т	S						L							<u> </u>				<u> </u>
88	T	I									<u> </u>	L		<u> </u>					
95	N	C			C						C	C		<u> </u>	<u> </u>				
97	P	S		S						S	<u> </u>	L	S	S	<u> </u>	ļ	<u> </u>		<u> </u>
107	S	D			D				D	L			<u> </u>		<u> </u>	ļ	D		
123	S	A				<u> </u>	A	A			<u> </u>	<u> </u>	L	<u> </u>		<u> </u>	A		├ ─
132	I	V	V		<u> </u>		V		V		<u> </u>	<u> </u>	<u> </u>	<u> </u>				<u> </u>	ļ
138	E	Α	Ĺ. <u></u> .	<u> </u>			Α		A		L		<u> </u>		A				├
145	M	F				F		<u> </u>	F			<u> </u>		└	<u> </u>	<u> </u>	F	F	F
151	Y	Α		Α			L	<u> </u>			A		ļ		Ь	ļ		L	A
167	V	I			I	<u> </u>			<u> </u>		I	<u> </u>	ļ	ļ	└ ─	<u> </u>	<u> </u>		ļ
180	L	I			I			ļ	<u> </u>	I	ļ	<u> </u>	ļ	ļ. <u>.</u>	ļ		<u> </u>	I	
194	Y	S	<u>L</u> _		L		S	<u> </u>	<u> </u>	S		<u> </u>	<u> </u>	S	├	 	<u> </u>	ļ	
199	A	S		S	<u> </u>			<u> </u>		S	 	<u> </u>	 -	 	 	-	├	<u> </u>	S
208	K	Н	<u> </u>	L		<u> </u>	H	H			<u> </u>		↓ _		_	ļ	-	<u> </u>	 -
209	N	K		<u> </u>	<u> </u>	<u> </u>	L_	ļ	<u> </u>		<u> </u>	-	↓	-	ļ			<u> </u>	├──
233	S	N		<u> </u>	Ļ			 		├ ─		ļ	ļ	↓	├	 ,,	<u> </u>	├ ─	 -
236	A	V	V	V	<u> </u>	ļ <u>.</u>	<u> </u>	V	ļ	<u> </u>		 		ļ	├	V	₩		
237	R	N	<u></u>	<u> </u>	<u> </u>	N	<u> </u>	ļ	<u> </u>		N	<u> </u>	 	-	├ ─		 	-	├
265	P	S	S	ļ		S	ļ	<u> </u>	↓	S	↓					 	<u> </u>	1	
267	V	I	I	<u> </u>	ļ	I	<u> </u>	I			 -	├	├ ──	├	 	-		I	┼
273	S	T	<u> </u>	<u> </u>	 	<u> </u>	T_	↓ .	 	ļ	T	┼—	 	+-	┼			 	
293	G	A	ļ	<u> </u>	A	<u> </u>	 	A	 	ļ	 	 	 -	-	┼		+	-	┼
299	L	C		C	 	<u> </u>	<u> </u>	C	 	1	<u> </u>	├ ──	l c	C	╂	-	+-	K	K
310	I	K	1	 	K	 	 	 	 	K	B	↓	┼	 -	╂	 	 	K	R
332	K	R	R	<u> </u>	<u> </u>	R	——	4—	 	 	R	-	↓	├ ──	╂	+	-	₩-	+ K
337	S	N	<u> </u>	 	<u> </u>	N	 	 	N			┼	-	 	┼─	┼──	-	s	┼──
355	P	S	S	S	 -	-	 	┼	S	-	┼	+	+	┼	-		+-	13	┼
362	L	M	 	-	↓	-	<u> </u>		 	├	 	-	+	┼	+-	┼─	\vdash	+-	+
363	A	V	↓	 		┼	 	┼	 	 		 	┼	 	 	+	 	+-	+
369	A	V		Ь	<u>L</u>	<u> </u>	<u> </u>		Щ.	<u> </u>	ــــــــــــــــــــــــــــــــــــــ		<u> </u>	<u> </u>	ــــــــــــــــــــــــــــــــــــــ		1	<u> </u>	ــــــــــــــــــــــــــــــــــــــ

Figure 16D

Sequen	ce cha	nges								$\overline{\mathbf{v}}$	arian	ts							
Position	WT	Mut	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
25	Y	Н																	
34	Α	S																	
48	K	E																	
50	D	N																	
55	N	S																	
63	T	S																	
88	T	I	I																
95	N	С																	
97	P	S				_													
107	S	D																	
123	S	Α																	
132	I	V																	
138	E	Α																	
145	M	F																	
151	Y	Α	A	A	Α	Α													
167	V	I	ī	I	I	I							I	I			I	I	I
180	L	I	I	I	I	I													
194	Y	S					S	S	S	S									
199	A	S					S	S	S	S									
208	K	H									H	Н							
209	N	K									K								
233	S	N											N						
236	Α	V																	
237	R	N													N	N	N	N	N
265	P	S															Ĺ		
267	V	I					I	I	I	I	I	I	I	I	I	I			1
273	S	T																L	
293	G	Α	·								Α	Α	Α	A	Α	Α	Α	Α	Α
299	L	С																	
310	I	K												L					
332	K	R									R	R	R	R	R	R	R	R	R
337	S	N											N	N					N
355	P_	S													S	S	S	S	S
362	L	M																	
363	A	V																	
369	Α	V																	

Figure 16E

Sequen	ce chan	ges	T						Var	iants									
Position	WT	Mut	65	66	67	69	70	71	72	73	74								
25	Y	Н	T																
34	A	S										Ĺ							
48	K	E						Е			<u> </u>		L						
50	D	N									<u>. </u>		L				_	\Box	
55	N	S						<u> </u>									\perp		
63	T	S							S	<u></u>			<u> </u>]	_	
88	T	I			<u> </u>							<u> </u>	<u> </u>				_		
95	N	С			<u> </u>								<u> </u>				_	_	
97	P	S				<u> </u>				<u> </u>	<u> </u>		<u> </u>]	
107	S	D								<u> </u>			<u> </u>	ļ			\sqcup		Ш
123	S	A				<u> </u>						<u> </u>					\square		
132	I	V				ļ					<u> </u>	<u> </u>	ļ	1_			\Box		\sqcup
138	E	A				<u> </u>	<u> </u>		<u> </u>		<u> </u>		ļ	<u> </u>	L.		igsqcup		\sqcup
145	M	F				L	<u> </u>	<u> </u>	<u> </u>	<u> </u>	<u> </u>	<u> </u>	<u> </u>	<u> </u>					
151	Y	_ A			<u></u>		<u> </u>	<u> </u>	<u> </u>	ļ	ļ	!	<u> </u>	1_	_				Ш
167	V	I	I	I		I	I	I	I	ļ	<u> </u>	ļ	<u> </u>	1_	_			<u> </u>	\vdash
180	L	I	<u> </u>	<u> </u>	<u> </u>	ļ	ļ	ļ	<u> </u>		<u> </u>	Ļ		 	_				igspace
194	Y	S			<u> </u>	<u> </u>	<u> </u>		<u> </u>		<u> </u>	<u> </u>	<u> </u>	╀-					\sqcup
199	A	S				<u> </u>	<u> </u>	1		ļ	ļ. <u></u>	<u> </u>		ļ		L-			
208	K	H			<u> </u>	<u> </u>	ļ	H	H	H	H		<u> </u>	-	_	L	_		
209	N	K	<u> </u>	L	<u> </u>		<u> </u>	<u> </u>	ļ	↓_	 	<u> </u>	1		<u> </u>	<u> </u>			<u> </u>
233	S	N	<u> </u>	<u> </u>		<u> </u>			ļ	<u> </u>	 _	<u> </u>	 	—	<u> </u>	ļ			<u> </u>
236	A	V	<u> </u>	<u> </u>	<u> </u>	<u> </u>	<u> </u>	<u> </u>	<u> </u>	 		↓ —	├ ─	+	L	_	\vdash		\vdash
237	R	N	<u> </u>	ļ	ļ	-		N	N	<u> </u>	 	ļ	↓	+	\vdash	 -	\vdash		<u> </u>
265	P	S	ļ	<u> </u>	<u> </u>	ļ. <u>.</u> _	 _	↓_	<u> </u>	 _	 -		 -		-	_			_
267	V	I	I	I	I	I	I	I	I	1	I	 	 	+	-	 —	\vdash		⊢-
273	S	T		<u> </u>	—	T	T	↓	 	 	 	 	₩	4-	 	 	\vdash	<u> </u>	
293	G	Α	A	<u> </u>	 	 		A	A	A	A	 		+	-	 			\vdash
299	L	С		 -	-	 	 	-	—	 	+-		┼		-	⊢	\vdash	-	
310	I	K	+_	 _ -	+_	 -	1-	 	<u> </u>	D	D	┼	₩-	+	-	-	\vdash	\vdash	-
332	K	R	R	R	R	R	R	R	R	R	R	├	+	4-	-	├-	<u> </u>	<u> </u>	\vdash
337	S	N	 	-	N	N	N	N	N	 	├	┼	+-	+-	┢	-	-	_	
355	P	S	1		S		-	S	S	 	┼	├—	1	+	-	 	-	 	-
362	L	M	-	1		↓	\vdash	↓	 	├ ──	₩-	-	┼-	+	-	├—	\vdash		
363	A	V	-	├		 	┼	+	 	1	177	-		+-	┼	 	\vdash	-	-
369_	A	V		<u></u>			Т		ــــــــــــــــــــــــــــــــــــــ	<u>V</u>	V	<u> </u>	Ь	ــــــــــــــــــــــــــــــــــــــ	1	<u> </u>		<u> </u>	Щ.

Figure 16F

Variant	y1	y 2	y4	у5	у6	у7
wt	0.7526	0.8774	0.7477	1.1850	0.6604	2
wt	1.2316	1.0877	1.2523	0.8150	1.3396	2
wt	1.0822	0.9082	1.0894	1.0850	0.9829	2
wt	0.8904	1.1423	0.9106	0.9158	1.0171	2
10	0.0263	1.7208	0.1682	-0.0125	0.0453	6
12	0.2211	0.1878	0.4486	0.2320	0.0415	2
13	0.0158	1.9119	0.2430	-0.0376	0.0302	1
14	0.0158	2.3899	0.3364	0.0251	0.0377	6
15	1.6789	0.0135	2.3738	1.6176	0.0226	
15	1.3945	0.4917	1.6260	1.2690	0.6857	
19	0.9000	0.9476	1.0280	1.0219	0.8528	1
19	0.6932	1.0442	0.6667	0.8143	0.7238	1
20	1.2737	0.0593	1.5327	1.5172	0.0755	0
20	0.5507	0.0484	0.5203	0.6472	0.0267	0
21	0.1632	0.9251	0.0935	0.1881	0.1509	0
22	0.1947	0.3294	0.1869	0.2884	0.0642	2
23	1.8053	0.0878	3.0280	2.0000	0.1585	3
23	1.6932	0.0900	2.0163	1.6709	0.1524	3
24	0.0579	0.9777	0.0374	0.0627	0.0566	4
25	0.3421	1.4891	0.6168	0.4514	0.5094	6
26	0.0053	10.7547	0.2056	0.0094	0.0566	2
26	0.0521	0.4391	0.0650	0.0127	0.0229	2
27	0.3474	1.3905	0.3178	0.3793	0.4830	1
29	1.4263	0.0079	1.6822	1.6144	0.0113	4
29	1.2740	0.0150	1.7398	1.3431	0.0190	4
30	0.0316	0.9560	0.0935	-0.0251	0.0302	8
31	0.0421	1.2547	0.1121	0.0502	0.0528	6
32	0.7316	1,2792	0.6916	1.0063	0.9358	4
33	0.3263	1.3530	0.5794	0.5235	0.4415	4
35	1.0737	0.1546	1.7009	1.4451	0.1660	1
36	0.0421	0.9858	0.2617	0.0752	0.0415	2
37	0.0316	0.9560	0.0187	-0.0094	0.0302	2
38	0.0053	9.3208	-0.0748	-0.0157	0.0491	0
39	0.2158	1.2416	0.2430	0.3730	0.2679	1
40	1.6737	1.5444	2.5794	2.0031	2.5849	2
40	0.9342	1.4557	0.9593	0.9666	1.3600	2
41	0.9421	1.8906	1.1402	1.2539	1.7811	2
42	0.0474	1.3543	0.0935	0.0784	0.0642	0
43	0.4105	0.1287	0.5794	0.6364	0.0528	4
43	1.0466	0.0109	0.9919	0.6113	0.0114	4
46	0.4466	1.0919	0.3089	0.5245	0.4876	0.5
47	0.6575	0.7763	0.6016	0.8143	0.5105	7
48	0.9370	0.8253	0.9919	1.1168	0.7733	4
51	0.0219	1.5643	-0.0488	0.0127	0.0343	1
55	1.0329	1.4901	1.0569	1.2986	1.5390	5

Figure 17A

Variant	yl	y2	y4	у5	у6	у7
56	1.3178	1.3124	1.5447	1.3198	1.7295	2
57	1.3123	0.9957	1.3496	1.2986	1.3067	3
58	0.9699	0.4635	0.8943	1.0237	0.4495	2
59	0.5260	0.0435	0.2927	0.5753	0.0229	0.5
60	0.5863	0.0325	0.3740	0.6578	0.0190	0.5
61	0.8548	0.0089	0.9268	0.9137	0.0076	0.5
62	0.3041	0.0752	0.2276	0.3574	0.0229	0.5
63	0.9370	0.0447	0.9919	0.9941	0.0419	0.5
65	0.3699	0.3193	0.0976	0.3955	0.1181	2
66	0.9096	0.5445	0.7480	0.9835	0.4952	2
67	0.2932	0.0520	0.1626	0.3194	0.0152	2
69	0.2301	0.2980	0.1951	0.1713	0.0686	2
70	0.5342	0.3066	0.2927	0.6028	0.1638	3
71	0.2411	0.3002	0.2114	0.2686	0.0724	0.5
72	0.4466	0.0427	0.2276	0.4611	0.0190	1
73	0.2219	0.7725	0.1138	0.2390	0.1714	4
74	0.7233	1.1113	0.4715	0.8164	0.8038	4

Figure 17B

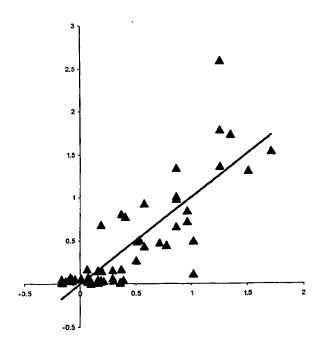


Figure 18

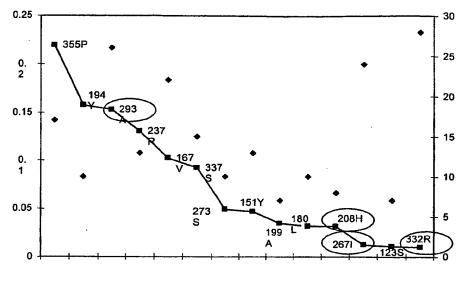


Figure 19

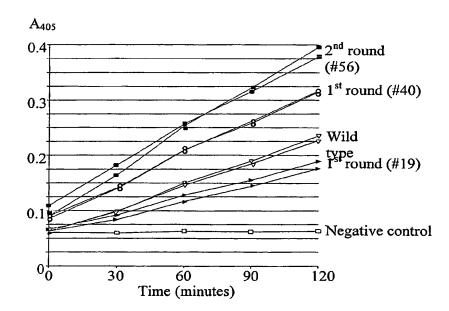


Figure 20

Variation position	Casein hydrolysis	Thermal tolerance	AAPL-p-NA pH7.0
	(y7)	(y6)	(y1)
107	<u>D</u>	S	S
123	<u>A</u>	S	S
151	<u>A</u>	Y	<u>A</u>
167	Ī	V	\overline{v}
180	<u>Ī</u>	L	<u>I</u>
194	<u>s</u>	Y	$ar{ extbf{Y}}$
199	<u>S</u> S K	Α	Α
208	K	<u>H</u>	K
267	V	Ī	V
273	<u>T</u>	Š	S
293	$\bar{\mathbf{G}}$	A	<u>A</u>
332	<u>R</u>	$\overline{\overline{\mathbf{R}}}$	$\frac{\overline{R}}{R}$

Figure 21

Align Im-1 heavy chain sequence with human germline ig heavy chain sequences from VBase using ClustalW.

A: Substitutions set

RULE 1a:

- Enumerate and classify the substitutions into 2 categories. (i) Substitutions found in the framework region and (ii) substitutions found in the CDR.
- Consider only these susbtitutions (ie RULE 1a is a filter)

B: Substitutions from human germline sequences

Reconstruct phylogenetic tree

RULE 1b:

 Calculate evolutionary proximity of the closest homolog in which each substitution occurs (EP)

RULE 2b:

- Calculate site heterogeneity at each substitution position (SH)
- Calculate entropy at each substitution position (SE) RULE 4b:
- Calculate number of times a substitution is seen at a position in the set of homologs (SN)

C: Substitutions from substitution matrices

RULE 1c:

Calculate favorability of each substitution using a PAM100 matrix (SM).

D: Score

 $Score_{FW} = f(EP) x f(SH) x f(SE) x f(SN) x f(SM)$

 $Score_{CDR} = f(SE) \times f(SN) \times f(SM)$

Figure 22

Align RSV-19 heavy chain sequence with human germline ig heavy chain sequences from VBase using ClustalW.

A: Substitutions set

RULE 1a:

- Enumerate and classify the substitutions into 2 categories. (i) Substitutions found in the framework region and (li) substitutions found in the CDR.
- Consider only these susbtitutions (ie RULE 1a is a filter)

B: Substitutions from human germline sequences

Reconstruct phylogenetic tree

RULE 1b:

 Calculate evolutionary proximity of the closest homolog in which each substitution occurs (EP)

RULF 2b

- Calculate site heterogeneity at each substitution position (SH) RULE 3b:
- Calculate entropy at each substitution position (SE)

RULE 4b:

 Calculate number of times a substitution is seen at a position in the set of homologs (SN)

C: Substitutions from substitution matrices

RULE 1c:

Calculate favorability of each substitution using a PAM100 matrix (SM).

D: Score

 $Score_{FW} = f(EP) x f(SH) x f(SE) x f(SN) x f(SM)$

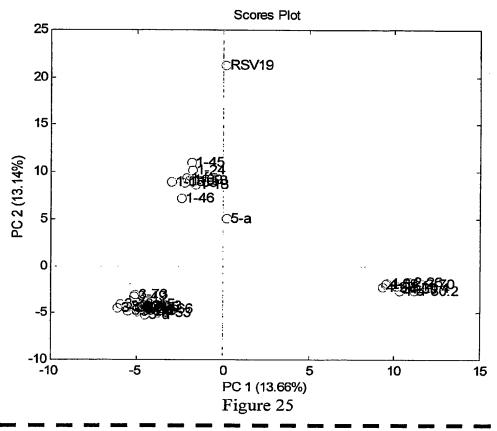
 $Score_{CDR} = f'(SE) \times f'(SN) \times f'(SM)$

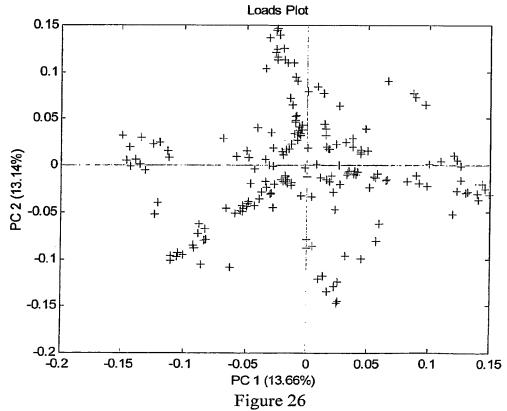
Figure 23

```
Locus-1-03
Locus-1-08
Locus-1-18
Locus-1-24
Locus-1-45
Locus-1-46
Locus-1-58
Locus-1-69
Locus-1-e
Locus-1-f
Locus-2-26
Locus-2-70
Locus-3-09
Locus-3-11
Locus-3-13
Locus-3-15
Locus-3-20
Locus-3-21
Locus-3-23
Locus-3-30
Locus-3-30.3
Locus-3-30.5
Locus-3-33
Locus-3-43
Locus-3-48
Locus-3-49
Locus-3-53
Locus-3-64
Locus-3-66
Locus-3-72
Locus-3-73
Locus-3-74
Locus-3-d
Locus-4-28
Locus-4-30.1
Locus-4-30.2
Locus-4-30.4
Locus-4-31
Locus-4-34
Locus-4-39
Locus-4-59
Locus-4-61
Locus-4-b
```

Figure 24

Locus-5-a





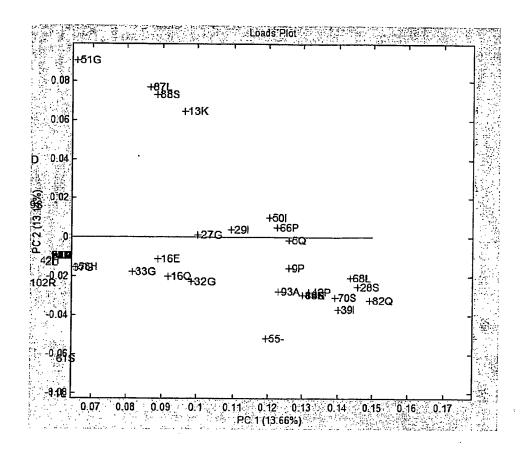


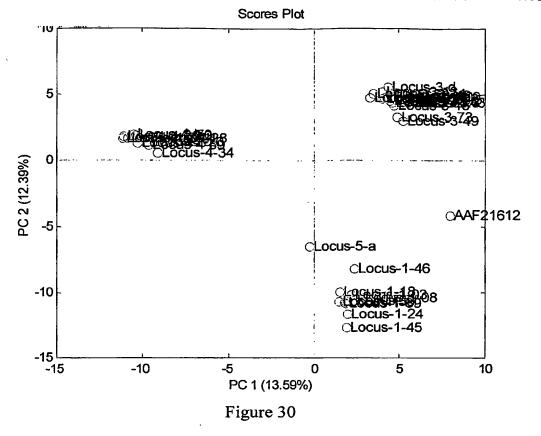
Figure 27

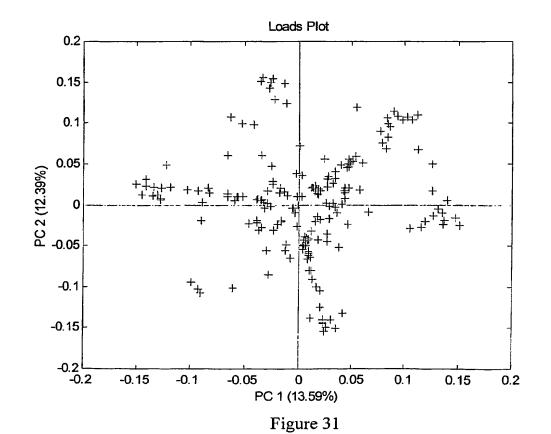
Substitutions	PC1 loads
27G	0.1001
291	0.1094
501	0.1204
66P	0.1226
93A	0.1229
5Q	0.1259
9P	0.126
15S	0.13
198	0.13
65N	0.13
83F	0.13
84S	0.13
86K	0.13
42P	0.1317
70S	0.1392

Figure 28

Locus-1-03 Locus-1-08 Locus-1-18 Locus-1-24 Locus-1-45 Locus-1-46 Locus-1-58 Locus-1-69 Locus-1-e Locus-1-f Locus-2-26 Locus-2-70 Locus-3-09 Locus-3-11 Locus-3-13 Locus-3-15 Locus-3-20 Locus-3-21 Locus-3-23 Locus-3-30 Locus-3-30.3 Locus-3-30.5 Locus-3-33 Locus-3-43 Locus-3-48 Locus-3-49 Locus-3-53 Locus-3-64 Locus-3-66 Locus-3-72 Locus-3-73 Locus-3-74 Locus-3-d Locus-4-28 Locus-4-30.1 Locus-4-30.2 Locus-4-30.4 Locus-4-31 Locus-4-34 Locus-4-39 Locus-4-59 Locus-4-61 Locus-4-b Locus-5-a

Figure 29





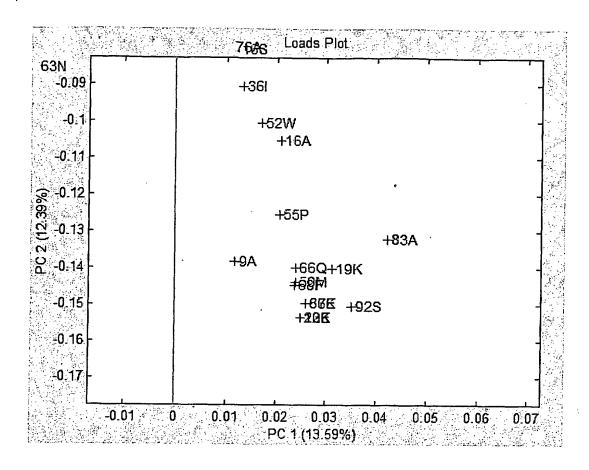


Figure 32

Substitutions	PC1 Loads
82Q	-0.1507
68L	-0.1452
391	-0.1421
708	-0.1421
28\$	-0.135
5Q	-0.1344
158	-0.1344
198	-0.1344
65N	-0.1344
83F	-0.1344
848	-0.1344
86K	-0.1344
66P	-0.1303
501	-0.1281
9P	-0.128

Figure 33